Applied Regression Analysis

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics, University of Washington

Session 7

Applied Regression Analysis, June, 2003 — 1

Adjusting for Covariates: Confounding, Precision, Effect Modification

Applied Regression Analysis, June, 2003 — 2

Confounding, Precision, Effect Modification

- Discriminating between confounding, precision, and effect modifying variables
  - Is the estimate of association between response and the predictor of interest the same in all strata?
    - Effect modifier: NO; Confounder, precision: YES
  - Is the third variable causally associated with the response after adjusting for the predictor of interest?
    - Confounder, precision: YES
  - Is the third variable associated with the predictor of interest?
    - Confounder: YES; Precision: NO

Applied Regression Analysis, June, 2003 — 3

Interpretation of Regression Parameters

- Difference in interpretation of slopes

Unadjusted Model : $E[Y_i \mid X_i] = \beta_0 + \beta_1 \times X_i$

- β_1 = Diff in mean Y for groups differing by 1 unit in X
  - (The distribution of W might differ across groups being compared)

Adjusted Model : $E[Y_i \mid X_i, W_i] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$

- γ_1 = Diff in mean Y for groups differing by 1 unit in X, but agreeing in their values of W

Relationship Between Models

- Relationship between the adjusted and unadjusted slopes
  - The slope of the unadjusted model will tend to be

$$\beta_1 = \gamma_1 + r_{XW} \frac{\sigma_W}{\sigma_X} \gamma_2$$

  - Hence, adjusted and unadjusted slopes for X are estimating the same quantity only if
    - $r_{XW} = 0$ (X and W are uncorrelated), OR
    - $\gamma_2 = 0$ (there is no association between W and Y after adjusting for X)

Relationship Between Models

- Relationship between the precision of the adjusted and unadjusted models

Unadjusted Model $\left[se(\hat{\beta}_1)\right]^2 = \dfrac{Var(Y \mid X)}{nVar(X)}$

Adjusted Model $\left[se(\hat{\gamma}_1)\right]^2 = \dfrac{Var(Y \mid X, W)}{nVar(X)(1 - r_{XW}^2)}$

$$Var(Y \mid X) = \gamma_2^2 Var(W \mid X) + Var(Y \mid X, W)$$

Example: Unadjusted Analysis (Case 1: A Precision Variable)

Fruit sizes by treatment group

	Fert	Sham	Diff
	3.7, 12.5,	41.6, 10.3,	
	13.7, 44.2,	0.9, 40.5,	
	43.8, 43.5,	9.8, 10.2,	
	4.3, 14.0,	11.1, 1.1,	
	4.6, 43.9,	39.9, 1.3,	
	13.8, 4.2	40.7, 1.4	
Mean	20.5	17.4	3.1
SD	17.7	17.6	

Example: Adjusted Analysis (Case 1: A Precision Variable)

Fruit sizes by treatment group and type of fruit

	Fert	Sham	Diff
Berry	3.7, 4.3,	0.9, 1.1,	
	4.6, 4.2	1.3, 1.4	
Mean(SD)	4.2 (0.37)	1.2 (0.22)	3.0
Apple	13.8, 12.5,	9.8, 10.2,	
	13.7, 14.0,	11.1, 10.3,	
Mean(SD)	13.5 (0.68)	10.4 (0.54)	3.1
Melon	44.2, 43.8,	41.6, 40.5,	
	43.5, 43.9	39.9, 40.7	
Mean(SD)	43.8 (0.29)	40.7 (0.70)	3.1

Example: Unadjusted Analysis (Case 2: A Confounder)

Fruit sizes by treatment group

	Fert	Sham	Diff
	3.7, 12.5,	41.6, 10.3,	
	13.7, 44.2,	0.9, 40.5,	
	3.8, 43.5,	9.8, 10.2,	
	4.3, 14.0,	11.1, 1.1,	
	4.6, 43.9,	39.9, 41.3,	
	13.8, 4.2	40.7, 1.4	
Mean	17.2	20.7	-3.5
SD	16.6	18.1	

Example: Adjusted Analysis (Case 2: A Confounder)

Fruit sizes by treatment group and type of fruit

	Fert	Sham	Diff
Berry	3.7, 4.3, 3.8, 4.6, 4.2	0.9, 1.1, 1.4	
Mean(SD)	4.1 (0.37)	1.1 (0.25)	3.0
Apple	13.8, 12.5, 13.7, 14.0,	9.8, 10.2, 11.1, 10.3,	
Mean(SD)	13.5 (0.68)	10.4 (0.54)	3.1
Melon	44.2, 43.5, 43.9	41.6, 40.5, 41.3,39.9,40.7	
Mean(SD)	43.9 (0.35)	40.8 (0.67)	3.1

Applied Regression Analysis, June, 2003 10

Example: Unadjusted Analysis (Case 3: An Effect Modifier)

Fruit sizes by treatment group

	Fert	Sham	Diff
	3.7, 12.5, 13.7, 44.2, 43.8, 43.5, 4.3, 14.0, 4.6, 43.9, 13.8, 4.2	45.6, 10.3, 0.9, 44.5, 9.8, 10.2, 11.1, 1.1, 43.9, 1.3, 44.7, 1.4	
Mean	20.5	18.7	1.8
SD	17.7	19.6	

Applied Regression Analysis, June, 2003 11

Example: Adjusted Analysis (Case 3: An Effect Modifier)

Fruit sizes by treatment group and type of fruit

	Fert	Sham	Diff
Berry	3.7, 4.3, 4.6, 4.2	0.9, 1.1, 1.3, 1.4	
Mean(SD)	4.2 (0.37)	1.2 (0.22)	3.0
Apple	13.8, 12.5, 13.7, 14.0,	9.8, 10.2, 11.1, 10.3,	
Mean(SD)	13.5 (0.68)	10.4 (0.54)	3.1
Melon	44.2, 43.8, 43.5, 43.9	45.6, 44.5, 43.9, 44.7	
Mean(SD)	43.8 (0.29)	44.7 (0.70)	-0.8

Applied Regression Analysis, June, 2003 12

FEV Example

Scientific Question

- Association between smoking and lung function in children
  - Longterm smoking is associated with lower lung function
  - Are similar effects observed in short term smoking in children?

Causal Pathway of Interest

- We are interested in whether smoking will cause a decrease in lung function as measured by FEV

Causation versus Association

- Statistical analyses, however, can only detect associations between smoking and FEV

  - In a randomized trial, we could infer from the design that any association must be causal
  - In an observational study, we must try to isolate causal pathways of interest by adjusting for covariates

Study Design

- Observational study
  - Measurements on 654 healthy children
    - Predictor of interest: Self-reported smoking
    - Response: FEV
    - Additional covariates
      - Effect modifiers
      - Potential confounders
      - Precision variables

Additional Covariates: Effect Modifiers

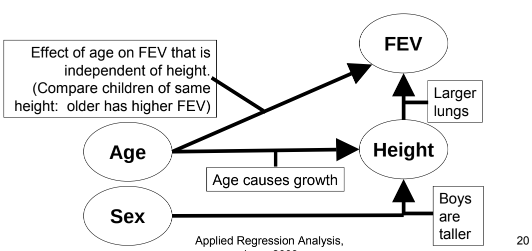
- There are no covariates currently of scientific interest for their potential for effect modification
  - First things first
    - Not generally advisable to go looking for different effects of smoking in subgroups before we have established that an effect exists overall
      - (We may sometimes delay discovery of important facts, but most times this seems the logical strategy)

Additional Covariates: Confounders

- Think about potential confounders
  - Necessary requirements for confounders
    - Associated causally with response
    - Associated with predictor of interest in sample
  - Prior to looking at data, we cannot be sure of the second criterion
    - But, clearly, any strong predictor of the response has the potential to be a confounder
      - So first consider known predictors of response
    - Furthermore, in an observational study, known associations in the population will likely also be in the sample

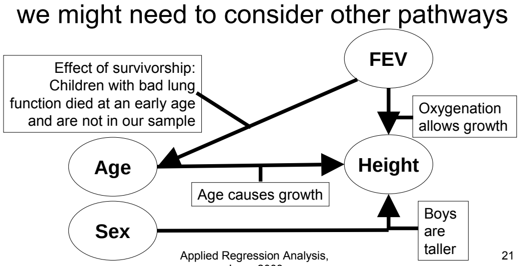
Predictors of FEV

- "Known" predictors of FEV



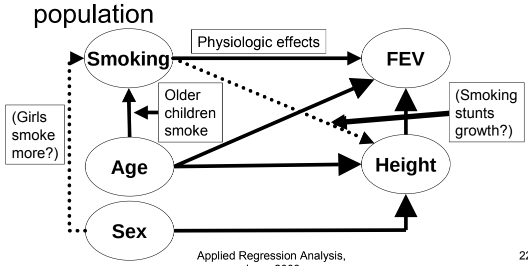
An Aside: What is "Known"?

- In an observational, cross-sectional study, we might need to consider other pathways



Associations with Smoking

- "Known" associations with smoking in the population

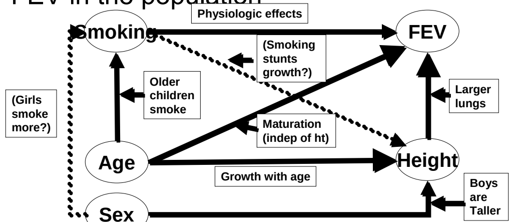


Adjusting for Potential Confounders

- Investigating the effect of smoking on FEV in children
  - We are scientifically interested in the possibility that smoking might cause decreased FEV
  - We are not scientifically interested in showing that FEV status might influence smoking behavior
    - (Of course, this is one possible explanation of an observed association, and so we must try to rule this out)

Associations with Smoking, FEV

- "Known" associations with smoking and FEV in the population



Pathways Tested in Unadjusted Analysis

- Comparing nonsmokers to smokers in observational study